Unscented Message Passing for Arbitrary Continuous Variables in Bayesian Networks

Wei Sun[*] [†] and **Kuo-Chu Chang**

Department of Systems Engineering and Operations Research
George Mason University
Fairfax, VA 22032

Abstract

Since Bayesian network (BN) was introduced in the field of artificial intelligence in 1980s, a number of inference algorithms have been developed for probabilistic reasoning. However, when continuous variables are present in Bayesian networks, their dependence relationships could be nonlinear and their probability distributions could be arbitrary. So far no efficient inference algorithm could deal with this case except Monte Carlo simulation methods such as Likelihood Weighting. But with unlikely evidence, simulation methods could be very slow to converge. In this paper, we propose an efficient approximate inference algorithm called Unscented Message Passing (UMP-BN) for Bayesian networks with arbitrary continuous variables. UMP-BN combines unscented transformation — a deterministic sampling method, and Pearl's message passing algorithm to provide the estimates of the first two moments of the posterior distributions. We test this algorithm with several networks including the ones with nonlinear and/or non-Gaussian variables. The numerical experiments show that UMP-BN converges very fast and produces promising results.

Introduction

Pearl's message passing algorithm (Pearl 1988) is the first exact inference algorithm developed originally for polytree discrete Bayesian networks. Applying Pearl's algorithm in the network with loops returns approximate answers and this method is called loopy belief propagation. Due to its simplicity of implementation and its good performance, loopy propagation has become very popular in recent years (Murphy, Weiss, & Jordan 1999; Weiss & Freeman 1999). In discrete case, messages are represented and manipulated by probability vectors and conditional probability tables (CPTs) which is relatively straightforward. But when applying Pearl's algorithm for continuous variables, it is much more complicated to represent and manipulate the messages. In this paper, we first propose to use the first two moments, mean and variance of a probability distribution, to represent the continuous message regardless of its specific distribution. Secondly, the integration of messages is treated using the concept of information fusion where messages are weighted by the inverse of their uncertainty (variances). Finally, to deal with the potentially nonlinear functional relationships between continuous variables and their parents, we propose to use unscented transformation (Julier & Uhlmann 1996; Julier 2002) to derive the corresponding messages. Unscented transformation uses a deterministic sampling scheme and can provide good approximations of the first two moments for the continuous variable undergone nonlinear transformation.

In our algorithm, unscented transformation plays a key role for computing messages. Specifically, we use it to compute the π message of a node and the λ message from a node sending to its parent since both computations involve the conditional probability distribution (CPD) in which nonlinear functions may be specified.

In short, we denote the unscented transformation for X undergoing a function $Y = f(X)$ as the following:

$$(Y.mu, Y.cov) = UT(X \xrightarrow{f(X)} Y) \quad (1)$$

Unscented Message Passing Algorithm

Given an arbitrary continuous Bayesian network, messages are represented by mean and variance of the continuous distribution for every node. Then the key problem becomes how to compute those messages and pass them to parents and children. Recall in Pearl's algorithm the conventional propagation equations are (Equation 4.49–4.53 on page 183 in (Pearl 1988)):

$$BEL(X) = \alpha\pi(X)\lambda(X) \quad (2)$$

$$\lambda(X) = \prod_{j=1}^{n} \lambda_{Y_j}(X) \quad (3)$$

$$\pi(X) = \sum_{\mathbf{T}} P(X|\,\mathbf{T}) \prod_{i=1}^{m} \pi_X(T_i) \quad (4)$$

$$\lambda_X(T_i) = \sum_{X} \lambda(X) \sum_{T_k:\, k\neq i} P(X|\mathbf{T}) \prod_{k\neq i} \pi_X(T_k) \quad (5)$$

$$\pi_{Y_j}(X) = \alpha \left[\prod_{k\neq j} \lambda_{Y_k}(X) \right] \pi(X) \quad (6)$$

[*]PhD candidate, wsun@gmu.edu, 703-993-1696

[†]10269 Braddock Road, Fairfax, VA 22032

where $\mathbf{T}(T_1, T_2, ..., T_n)$ are the parents of node X; $\mathbf{Y}(Y_1, Y_2, ..., Y_m)$ are the children of node X; $\lambda_X(T_i)$ is the λ message from X to its parent T_i; $\pi_{Y_j}(X)$ is the π message from X to its child Y_j; and α is a normalizing constant.

In continuous case, obviously, integral replaces summation in the above equations. Recall that $\pi(X) = P(X|e_X^+)$ and $\lambda(X) = P(e_X^-|X)$ where $\{e_X^+, e_X^-\}$ representing the evidence from the sub-networks "above" and "below" a node X respectively. In recursive Bayesian inference, π message represents prior information and λ message represents evidential support in the form of a likelihood function. Equations 2, 3, 6 are essentially the data fusion by a product of different messages. Under the Gaussian assumption, it is straightforward to fuse multiple estimates (Bar-Shalom, Li, & Kirubarajan 2001, page 94). Specifically, we have:

$$BEL(X) \begin{cases} cov = \left(\frac{1}{\pi(X).cov} + \frac{1}{\lambda(X).cov}\right)^{-1} \\ mu = cov\left[\frac{\pi(X).mu}{\pi(X).cov} + \frac{\lambda(X).mu}{\lambda(X).cov}\right] \end{cases} \tag{7}$$

$$\lambda(X) \begin{cases} cov = \left(\sum_{j=1}^{n} \frac{1}{\lambda_{Y_j}(X).cov}\right)^{-1} \\ mu = cov\left[\sum_{j=1}^{n} \frac{\lambda_{Y_j}(X).mu}{\lambda_{Y_j}(X).cov}\right] \end{cases} \tag{8}$$

$$\pi_{Y_j}(X) \begin{cases} cov = \left(\frac{1}{\pi(X).cov} + \sum_{k \neq j} \frac{1}{\lambda_{Y_k}(X).cov}\right)^{-1} \\ mu = cov\left[\frac{\pi(X).mu}{\pi(X).cov} + \sum_{k \neq j} \frac{\lambda_{Y_k}(X).mu}{\lambda_{Y_k}(X).cov}\right] \end{cases} \tag{9}$$

Equation (4) computes the π message for node X. Theoretically, this is equivalent to treating X as a functional transformation of $\mathbf{T}$ and the function is the one defined in CPD of X denoted as $h(X)$. Technically, we take $\mathbf{T}$ as a multivariate random variable with a mean vector and a covariance matrix; then by using unscented transformation, we obtain an estimate of mean and variance of X to serve as the π message for node X. In Equation (4), $\pi_X(T_i)$ is the π messages sending to X from its parent T_i, which is also represented by 'mean' and 'variance'. By combining all the incoming $\pi_X(T_i)$ messages, we can estimate the mean vector and covariance matrix of $\mathbf{T}$. Obviously, the simplest way is to view all parents as independent variables; then take their means to compose the mean vector, and put their variances at the diagonal positions to form a diagonal matrix as the covariance matrix.[1] With that, we can compute the π message of node X by

$$(\pi(X).mu, \pi(X).cov) = UT(\mathbf{T} \overset{h(X)}{\longrightarrow} X) \tag{10}$$

Similarly but a bit more complicated, Equation (5) computes the λ message sending to its parent from node X. Note here that we integrate out X and all of its parents except the one (T_i) we are sending λ message to. Theoretically, this is equivalent to regarding T_i as the functional transformation of X and $\mathbf{T}\backslash T_i$. It is necessary to mention that the function used here for transformation is the inverse function of the original one with T_i as the independent variable. We denote this inverse function as $v(X, \mathbf{T}\backslash T_i)$. To compute the message, we first augment X with $\mathbf{T}\backslash T_i$ to obtain a new multivariate random variable called $\mathbf{TX}$; then the mean vector and covariance matrix of $\mathbf{TX}$ can be estimated easily by $\lambda(X)$ and $\pi_X(T_k)(k \neq i)$. After applying unscented transformation to $\mathbf{TX}$ with the new inverse function $v(X, \mathbf{T}\backslash T_i)$, we can compute $\lambda_X(T_i)$ message as below,

$$(\lambda_X(T_i).mu, \lambda_X(T_i).cov) = UT(\mathbf{TX} \overset{v(X,\mathbf{T}\backslash T_i)}{\longrightarrow} T_i) \tag{11}$$

With Equations (7) to (11), we can now compute all messages for continuous variables. As you may notice, unscented transformation plays a key role here. This is why we name this algorithm Unscented Massing Passing for Bayesian Network (UMP-BN).

Numerical Experiments and Conclusion

To test the algorithm, we use two real-world networks each has multiple loops with different sizes. One is INCINERATOR, borrowed from (Lauritzen 1992) in which the author proposed the Junction Tree algorithm for conditional linear Gaussian network (CLG). Another one is ALARM, used in the paper (Beinlich *et al.* 1989) to compare various inference algorithms. Those networks are either CLGs or discrete networks originally. In our experiments, we leave the network structures unchanged but specify different CPDs including the ones with nonlinear functions. Numerical experiments show that UMP-BN performs very well under various scenarios.

[1]This is actually why loopy estimation is not exact. To improve the algorithm, we can estimate the covariance between all parents and pass them into the covariance matrix for $\mathbf{T}$.

References

Bar-Shalom, Y.; Li, X. R.; and Kirubarajan, T. 2001. *Estimation with Applications to Tracking and Navigation*. Wiley.

Beinlich, I.; Suermondt, G.; Chavez, R.; and Cooper, G. 1989. The alarm monitoring system: A case study with two probabilistic inference techniques for belief networks.

Julier, S. J., and Uhlmann, J. K. 1996. A general method for approximating non-linear transformations of probability distribution. Technical report, RRG, Dept. of Engineering Science, University of Oxford.

Julier, S. J. 2002. The scaled unscented transformation.

Lauritzen, S. 1992. Propagation of probabilities, means amd variances in mixed graphical association models. *JASA* 87(420):1098–1108.

Murphy, K.; Weiss, Y.; and Jordan, M. 1999. Loopy belief propagation for approximate inference:an empirical study.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kauffman, San Mateo.

Weiss, Y., and Freeman, W. T. 1999. Correctness of belief propagation in gaussian graphical models of arbitrary topology. Technical report, Dept. of Computer Science, University of California at Berkeley.